

A Logic Programming Approach to Integration Network Inference

Daniel Ritter

SAP AG, Technology Development – Process and Network Integration,
Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
`daniel.ritter@sap.com`

Abstract. The discovery, representation and reconstruction of (technical) integration networks from Network Mining (NM) raw data is a difficult problem for enterprises. This is due to large and complex IT landscapes within and across enterprise boundaries, heterogeneous technology stacks, and fragmented data. To remain competitive, visibility into the enterprise and partner IT networks on different, interrelated abstraction levels is desirable.

We present an approach to represent and reconstruct the integration networks from NM raw data using logic programming based on first-order logic. The raw data expressed as integration network model is represented as facts, on which rules are applied to reconstruct the network. We have built a system that is used to apply this approach to real-world enterprise landscapes and we report on our experience with this system.

Keywords: Datalog, Integration Networks, Knowledge Representation, Logic Programming, Network Inference, Network Mining

1 Introduction

Enterprises are highly connected to partners and even competitors as part of value chains consisting of business processes. The business document exchange is actually implemented by complex, underlying networks of application and middleware systems, called integration networks. To remain competitive enterprises have to adapt their business processes in a timely and flexible manner, which requires visibility and control over the integration network. However, currently information is locked into systems of an enterprise. To overcome this situation, a new discipline, called Network Mining (NM), strives to discover and extract raw data hidden within heterogeneous systems in complex enterprise landscapes [21,20]. The raw data implicitly contains information about the integration network, i.e. middleware and application. From that, our system reconstructs integration networks. For the system user, the resulting linked real-world data describing the "as-is" network can then be captured in e.g. network-centric BPMN models [19].

A generalized view of such a network is shown in Fig. 1. When looking at an enterprise landscape, the systems within the integration network can be classified into different categories based on the integration content and the role they

play. The classification provides insight into the capabilities and complexity of the network and allows to manage business processes, contextualized visualization and operation on the network. These categories span from applications with embedded integration or even mediation capabilities, like proxies, enterprise services, composite applications or applications with service adaptation (Categories I+II), over standalone Enterprise Service Bus (ESB) or middleware instances with flexible pipeline processing, e.g. mapping, routing and connectivity for legacy systems (Category III+IV), to Business to Business (B2B) gateways for cross-enterprise document exchange (Categories V+VI) and system management solutions, which allow to operate these systems, their software and lifecycle (Category VII).

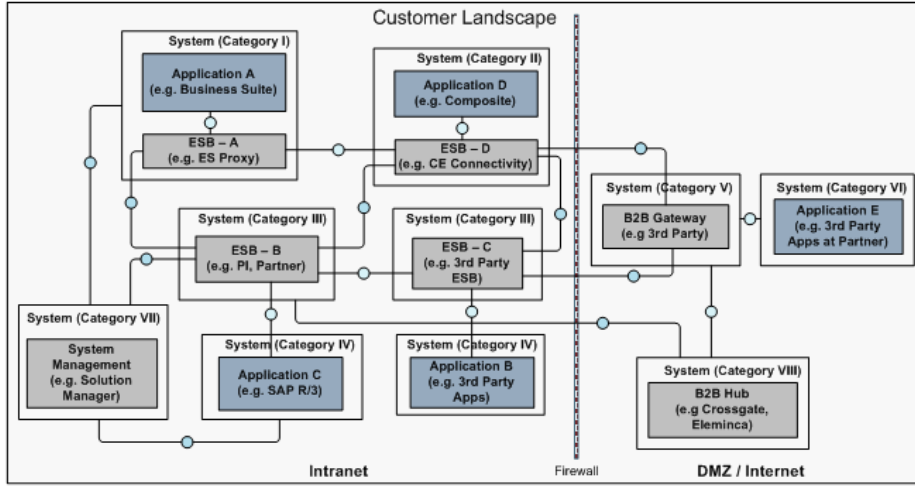


Fig. 1. Sample (technical) Integration Network showing logical systems as participants with embedded integration capabilities and standalone middlewares as well as B2B gateways

In this paper we present an approach to model and reconstruct integration networks from discovered raw data using logic programming, more precisely standard Datalog with recursion and stratified negation. We describe how information in form of NM raw data can be represented independent of their original domain in a Network Integration Model (NIM) and how user facts can be added. We have chosen Datalog to represent this model, which we use to develop Datalog programs (i.e. a finite set of Datalog rules) that express the network. That means identifying entity equivalences, computing edges and semantic references as well as dealing with user input. We validated our approach on simulated integration network data and report our experience with the network inference Datalog system in real-world enterprise networks as well as possible extensions.

In Section 2 we describe the problem domain and state on design principles and decisions in Section 3. Section 4 defines the NIM and Section 5 introduces the inference algorithm. Section 6 shows experimental results and states on ex-

periences. Section 7 concludes with related work, before we draw conclusions and outline future research in Section 8.

2 Motivation

Our premise is that relevant data for computing integration networks is hidden in enterprise system landscapes. However, for that it has to be discovered by NM from mostly disjoint domains in different formats with different meaning [20]. The integration networks derived from the discovered information consist of nodes and edges on different abstraction levels.

The basic entities of the integration network are logical systems (e.g. tenants, applications, integration middleware) and message flows, which are either direct connectivity or mediated communication/ integration. The actual information about these entities as well as their semantics are discovered by Network Mining (NM) systems [20]. However, the discovered raw data is domain-specific and needs to be translated into a domain independent model for network inference, while preserving its semantics. The definition of a Network Integration Model (NIM) is the basis for applying network inference algorithms. Since the raw data comes from disjoint domains, in different formats with different semantics, inference algorithms have to deal with possibly duplicate, fragmented, uncertain or incorrect information while computing the network. Fig. 2 schematically shows some of these challenges. For instance, entity equivalences have to be identified and handled. Direct and transitive edges have to be calculated and semantic relations between nodes have to be inferred. Fig. 2(a) shows systems SX_1 and SX_2 discovered from domain X exchanging messages over middleware system MWX_1 , and systems SY_1 and SY_2 discovered from domain Y exchanging messages over middleware system MWY_1 . Here, SX_2 and SY_2 denote the same system, as well as MWX_1 and SY_1 are equivalent. Based on the inferred equivalences, the nodes are partitioned as equivalence relations Eq , i.e. $Eq(MWX_1, SY_1)$ and $Eq(SX_2, SY_2)$, and the edges are computed accordingly (see Fig. 2(b)). Systems or applications run on physical hosts, e.g. H_1 from discovery domain $Host$. The relationships between systems and hosts are not considered as edges but semantic references within the network. Hosts build the bridge to the related domain of system management networks, which are addressed by [18,11]. A new host CS_1 is added to the network as user knowledge on which SY_1 runs. When merging systems MWX_1 and SY_1 the semantic relation is preserved.

3 Design Principles and Decisions

The major design decisions taken were about finding a representation for an integration model and a language to express inference algorithms. We needed to select (1) an approach, which does not require to modify the system when changing the inference programs or the integration model, (2) a well-understood

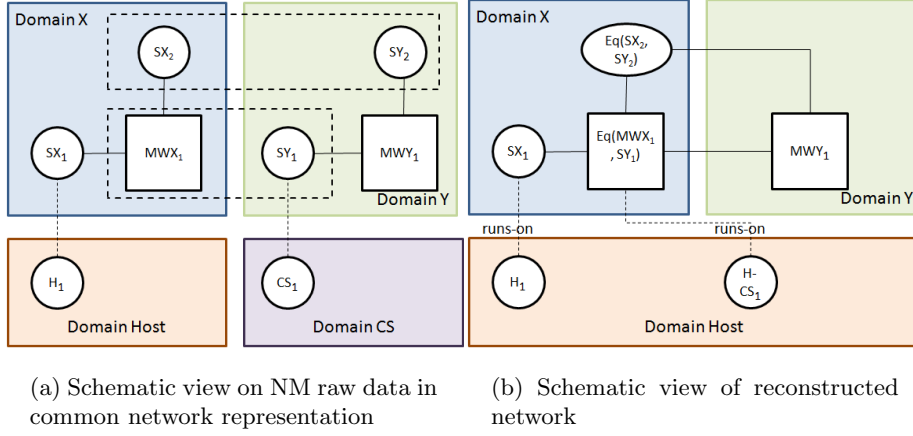


Fig. 2. Schematic view on the inference challenges

representation for information suitable for the inference approach, and (3) a sufficiently powerful inference technique, simple enough to be used by our customers and partners to define their own inference programs.

The necessity of (1) is derived from developing the inference programs in the early prototypes. The domain of the data and the scope of inference evolved - and it will continue to do so as more data sources are integrated and inference is refined. Hence the lifecycle of the data model and of the inference programs needs to be decoupled from that of the system. Since system landscapes and business networks for large enterprises are very complex and many implementations need customer-specific modifications or extensions both (2) and (3) are required. As the relational model is a foundation for most business applications and is thus well-understood by customers, it is a natural choice for (2). Consequently, we initially considered SQL and its imperative extensions to express inference programs. However, as network analysis and inference are expressed more naturally using recursive rules we moved towards logic programming languages like Prolog or Datalog, choosing Datalog for its simpler semantics.

4 The Integration Network (Inference) Model

The model for representing integration networks as virtual "as-is" enterprise landscape covers a representative intersection of entities from the enterprise integration middleware space [15]. Although this domain has many aspects, which are even differently treated in different system implementations, we identified a common, core meta-model, which we call Network Integration Model (NIM). The basic NIM entities relevant for the inference are introduced subsequently, while more entities might be explained later where necessary.

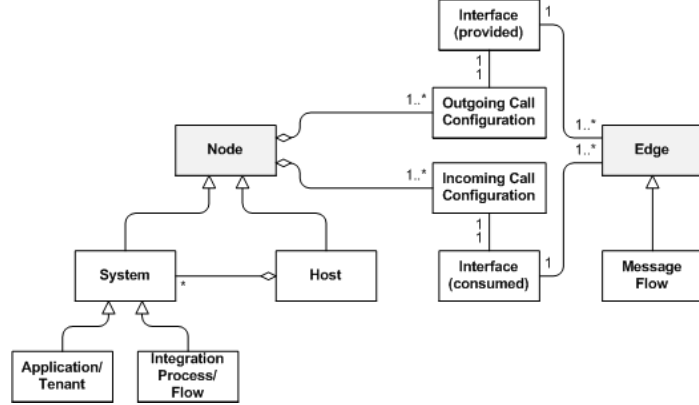


Fig. 3. The basic NIM entities and their relations

The base premise for defining an integration meta-model is to represent the actual physical hosts in the enterprise landscape as first class entities and then find the interfaces provided or called by them during message-based communication. Since most of the communication actually addresses logical entities like applications or tenants, called systems, running on the physical hosts, a *System* is considered a node of the network. That means, systems represent (business) application and integration logic. For the communication with other systems via messages the *MessageFlow* represents edges in the network. Technically, messages are exchanged over interfaces, *Interface*, and channels, containing e.g. service bindings and operations, which we represent as *IncomingConfiguration* and *OutgoingConfiguration*. The inbound and outbound configurations are considered separate entities, since they carry important information about the message flows, thus helping to reconstruct the network's edges. This notion can also be found in a common graph traversal algebra to set custom processors or actions when entering or leaving a node [23]. Fig. 3 shows the basic NIM entities and their relations.

5 The Network Inference Approach

The algorithm for computing integration networks consists of multiple steps, which have been identified for a parallel analysis allowing it to scale across large datasets of NM raw data. Since the information is represented in the NIM, the inference mechanism is independent of the specific integration and system domains. As discussed in Section 2, unique systems and hosts are identified by equivalence algorithms and semantic links between hosts and systems are computed (step 1). Based on that, incoming and outgoing configurations are identified (step 2) and then used to reconstruct message flows through building separate call graphs (steps 3,4) which are merged afterwards (step 5). Then message flows are linked with application and integration content (step 6) and

user knowledge is integrated. With user knowledge, the quality of the inference mechanism can be improved and information complemented or enriched. Within the inference programs, all user knowledge literals end with the "user" postfix, while discovered knowledge ends with "disc" (i.e. edb relation).

To formalize the network reconstruction, a logic programming approach is used, in which the algorithms are described by Datalog rules and the discovered raw data is a set of Datalog facts according to NIM. The different processes of adding newly discovered information and removing outdated is continuous. For that, each piece of discovered information is annotated with a timestamp. However, instead of removing outdated information that is referenced by higher layer information models as in [19], it is kept and marked outdated until it is not referenced anymore.

Step 1: Identify unique hosts and systems To identify hosts and systems uniquely through building equivalence classes, the single instances have to be identified. While hosts can be identified by e.g. host name, IP-address, the systems have no universally applicable identification scheme, thus they are usually identified using context dependent identifiers. For instance, the set of host identifiers can be an IP-address, the DNS name, and a host name. This information mainly comes from different, disjoint instances of system management software, mostly from IT service management [18] and virtualization systems [11]. All identifiers are contained in the equivalence class and any reference to one of them identifies the host. While these equivalence classes are not stable over time, it is quite likely that at least one of the elements of an equivalence class does not change if another one changes, thus making the identification more robust. That way, identity can be maintained over long periods of time in the presence of constant but gradual change. The raw facts from NM are *host_disc(host_id, URI)* and *system_disc(sys_id, URI)*, which relate a *host_id* or *sys_id* to an addressable URI. Relations like *same_host_disc(host_id1, host_id2)* and *same_sys_disc(sys_id1, sys_id2)* connect two host or system identifiers, e.g. which refer to the same physical host or logical system. The semantic relation *runs_on_disc(sys_id, host_id)* connects a system to the host that it runs on. For simplicity, homogenous clusters of machines are also considered as one host.

Listing 1.1. Host equivalence exploiting information about system landscape

```

same_sys(?sys_id1, ?sys_id2) :-
    same_sys_disc(?sys_id1, ?sys_id2).
same_sys(?sys_id1, ?sys_id2) :-
    same_sys(?sys_id1, ?sys_id3),
    same_sys(?sys_id3, ?sys_id2).

same_host(?host_id1, ?host_id2) :-
    runs_on_disc(?sys_id1, ?host_id1),
    runs_on_disc(?sys_id2, ?host_id2),
    same_sys(?sys_id1, ?sys_id2).

```

Based on that, rules for e.g. *same_sys* and *same_host* are used to infer equivalence classes that allow to write rules that exploit the information about system landscapes. For instance, more than one system can run on one physical host, but one system cannot run on more than one host, Listing 1.1.

Step 2: Determine Incoming and Outgoing Calls In current middleware route configurations, the senders of incoming calls to the system can be registered but are mostly unknown. On the other hand, components like the file adapter and the Apple Push Notification Service (APNS) always contain the sender system [15]. However, for outgoing calls from the sender system, e.g. via HTTP, SOAP, receiver or outgoing call configurations are needed to initiate the message flow to the receiver. This results in an outgoing and incoming call graph depicted in Fig. 4(a). The $incoming_disc(sys_id, URI)$ and $outgoing_disc(sys_id, URI)$ facts relate a sys_id to a URI of an incoming configuration or an outgoing configuration for the identified system.

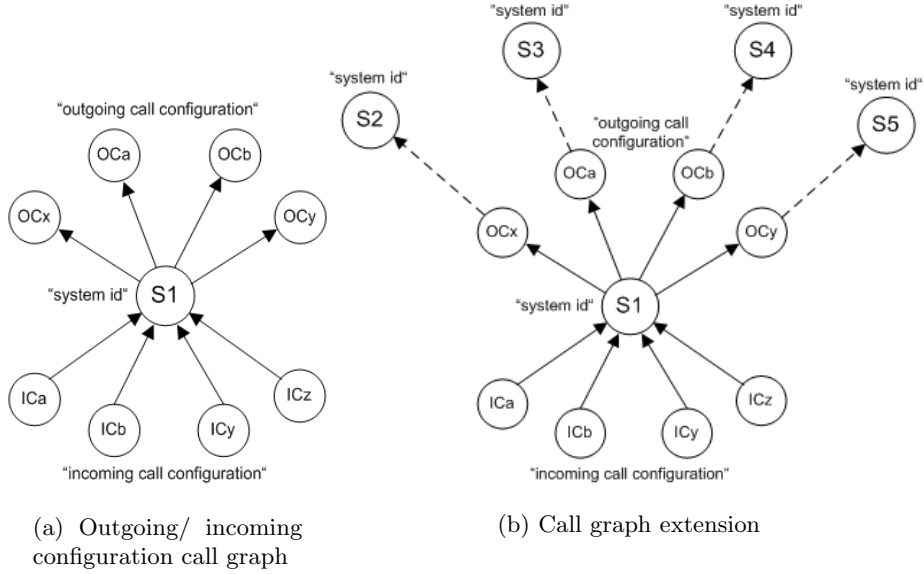


Fig. 4. Outgoing and incoming configuration call graphs

Step 3: Determine Message Flows based on Outgoing Calls Since outgoing calls are made to a particular endpoint, the corresponding call configurations contain an identifier for the receiving host or system. These identifiers can then be matched against the identifiers that were determined in step 1. If no identifiers are available, these call configurations are processed in step 4. To relate outgoing call configurations to receiver systems $recv_disc(URI, sys_id)$ relates a URI to an outgoing configuration to a sys_id that identifies a receiving system or similarly $recv_host_disc(URI, host_id)$ for hosts.

Listing 1.2. Message flow from outgoing configuration

```
msg_flow(?sys_id_snd, ?sys_id_recv) :-  
    outgoing_disc(?sys_id_snd, ?RCONF),  
    recv_disc(?RCONF, ?sys_id_recv).
```

Listing 1.3. Message flow for host configurations

```
msg_flow_host(?host_id_send, ?host_id_recv) :-  
    runs_on_disc(?sys_id_snd, ?host_id_send),  
    outgoing_disc(?sys_id_snd, ?RCONF),  
    recv_host_disc(?RCONF, ?host_id_recv).
```

Then *message_flow(sys_id_snd, sys_id_recv)* rules determine the message flows between systems (Listing 1.2) and *message_flow_host(host_id_send, host_id_recv)* between hosts (Listing 1.3). That results into a an extension of the call graph shown in Fig. 4(b), in which *S1* represents a system connected to other systems via incoming and outgoing configurations.

Step 4: Determine Message Flows based on Incoming Calls Similar to the previous step, incoming call configurations are identified. For that, *send_disc(URI, sys_id)* facts are related via URI to incoming configurations. Again, this results in an extension of the call graph.

Step 5: Merge Call Graphs for a System So far unique hosts and systems are identified and message flows are determined for a single system. Now, the identified incoming and outgoing call configurations from different systems are matched. This is done by matching compatible protocols, message types, etc. After new message flows are identified, the call graph is extended by the merged information (see Fig. 5). In case some incoming or outgoing call configurations do not match to already identified call configurations, they are kept in the model as "unlinked" configurations for matching new configurations.

Step 6: Link Message Flows to Application and Integration Content The outgoing and incoming call configurations with hosts and systems result in a view of the network. However, these message flows only conclude communication between hosts and systems. The outgoing and incoming call configurations also have a link to application and integration content deployed and running on the systems. This content refers to the particular process or integration steps that trigger outgoing calls or receive incoming calls. In other words, process models [1] and middleware routes [15], i.e. integration flow (IFlow) or integration process, give insight into the details within systems and hosts and could be used to correlate operational data to trace messages through middleware systems.

Listing 1.4. Identifying IFlows

```
iflow(?sys_id_snd, ?sys_id_recv, ?sys_id_mw, ?URI) :-  
    msg_flow_disc(?sys_id_snd, ?sys_id_mw, ?URI),  
    msg_flow_disc(?sys_id_mw, ?sys_id_recv, ?URI).
```

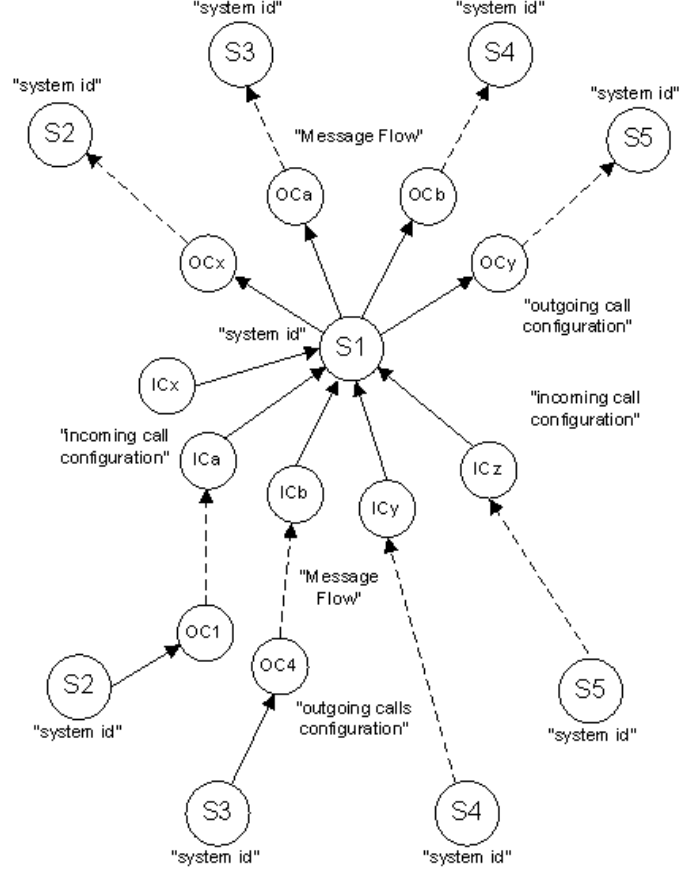



Fig. 5. Call graph extended by merged information from different systems

For instance, the IFlow $iflow(sendsys_id, recvsys_id, mwsys_id, URI)$ relates senders to receivers through a middleware system, which can be calculated e.g. through the rule in Listing 1.4.

6 Results and Experiences

For the evaluation of our approach, we used our Datalog system, which is a basic Datalog implementation in Java/OSGi based on [25], that allows to evaluate recursive rules and supports basic data types, comparisons and expressions in Datalog rules. The raw data comes from our Network Mining prototype, which discovers information in our testbed and transforms it to NIM Datalog facts. The testbed consists of two middleware systems, i.e. *HXP* and *H73*, of different releases for mediated communication, and have embedded IDoc and WebService capabilities for direct communication and a System Landscape Directory (SLD)

for system management information. This setup contains real-world conditions which we found in our customer landscapes, e.g. cross-middleware inference, combination of embedded and mediated communication, fragmented information registered in different domains.

The results of the experiments are shown, e.g. for systems and message flows in *HXP* in Table 1, 2 and for *H73* in Table 3, 4. The tables show two aspects of the system, namely the discovery and the inference quality. For the inference, the entries for systems and message flows as well as top-level connections are important. The discovery is mainly depicted by attribute entries for the network entities and show minor gaps in the discovery process., e.g. in the category "Correct System Attributes" (see Table 1). For the HXP-PI system, 12 nodes and

Table 1. Inference Result on NM Testbed (Nodes of HXP-PI System)

Category	Absolute Value	Percentage
<i>Found expected Systems</i>	12	100%
Correct System Attributes	35	64%
System Attributes with Limitation	20	36%

55 node attributes are expected (see Table 1). In total 13 top-level connections are expected which group 31 message flows (see Table 2). Furthermore, the top-level connection have 26 attributes, while the message flows have 372 attributes.

Table 2. Inference Result on NM Testbed (Edges of HXP-PI System)

Category	Absolute Value	Percentage
<i>Found Expected Top-Level Connection Groups</i>	13	100%
Correct Top-Level Connection Group Attributes	26	100%
<i>Found Expected MessageFlows</i>	31	100%
Correct MessageFlow Attributes	337	91%
MessageFlow Attributes with Limitation	34	9%

For the cross middleware systems and message flow inference, in total 18 unique, logical systems were inferred from 29 partially duplicate raw facts via equivalence determination (see Table 1 and 3) and 34 message flows have to be reconstructed and grouped to 17 top-level connections using incoming and outgoing call graph merge operations. For instance, logical system *HXP_105* from *PI-HXP* with runs-on host id *xxx2474* from SLD was found in the middleware configuration and SLD system information facts and merged into an equivalence class (see Table 5). At the same time, the corresponding message flows between *HXP_105* to *HXP_106* were reconstructed from PI configuration (conf.) and run-time (runt.) data connected to the system equivalence sets and merged into an

top-level connection group (see Table 6). The group consists of the message flow over sender interface *FlightSeatAvailQuery* to system *HXP_106*, which checks for free seats and is followed by a message to the same system over interface *BookOrderRequest* in case of a positive answer to the first query. If the booking order request was successful, system *HXP_106* answers over interface *FlightBookOrderConfirm* to confirm the request. No unexpected systems or message flows were found and the complete network structure was reconstructed correctly.

Similarly, the H73-PI system has 3 parties, i.e. B2B contexts, 6 expected nodes with 31 attributes (for Table 3), 4 top-level connections, grouping 6 flows (for Table 4), with 8 attributes on the top-level connections and 78 on the message flows.

Table 3. Inference Result on NM Testbed (Nodes of H73-PI System)

Category	Absolute Value	Percentage
<i>Found expected Parties</i>	3	100%
Found expected System	6	100%
Correct System Attributes	22	71%

Table 4. Inference Result on NM Testbed (Edges of H73-PI System)

Category	Absolute Value	Percentage
<i>Found Expected Top-Level Connection Groups</i>	4	100%
Correct Top-Level Connection Group Attributes	8	100%
Top-Level Connection Group Attributes with Limitation	0	0%
<i>Found Expected MessageFlows</i>	6	100%
Correct MessageFlow Attributes	75	96%
MessageFlow Attributes with Limitation	3	4%

The detailed inference results are only shown partially due to the mass of data discovered. Hence Table 5 shows an excerpt of the results of systems with the discovered description, the inferred host and the equivalence class denoted by "discovered system". Similarly, an excerpt of the inferred message flows are shown in Table 6. For that, the top-level connections, i.e. grouped message flows are listed with their message flows denoted by sender and receiver and the type of discovered facts from which the data came from (as "From"). In the excerpt, all message flows themselves build an equivalence class of same flows found in runtime logs (runt.) and configuration (config.).

Due to good results in our testbed, we applied the system to real-world customer landscapes as shown in Fig. 6. This real-world validation was very successful on both counts. Firstly, it proved that the auto-discovery and inference is indeed feasible and resulted in highly reliable results. Secondly, our system would

be quite helpful in the everyday work of an integration architect, consultant or integration developer, since it gives an overview of the complete integration network which is currently not possible within the integration middleware tools. The system reduces the effort to document integration scenarios substantially, in particular by a foreseen export of network details into PDF or office format. That helps to answer specific questions about the network, which are currently still impossible (or difficult) to achieve. For example, when combining configuration and runtime data it is possible to find connections that are not used any longer or were seldom used in a given period of time. Hence, one of the customers plan an upgrade project and with such a system a substantial migration time and effort will be saved.

7 Related Work

Our approach for integration network representation and inference is based on Datalog, which is a well-researched topic [12,25] that had its revival recently due to good parallelization capabilities, latest through the work of Hellerstein et al. [2,14]. Even in the enterprise analytics domain, Datalog was recently applied, mainly through work of [5,6,7]. However, these approaches address non-network inference domains for which they define extensions.

In terms of the meta-model for integration network, [23] represents closest known related work, in which a path algebra is defined that is used to traverse arbitrary graphs. Similarly we define nodes and edges with inbound and outbound connectors, however different in terms of meaning and usage.

Table 5. Excerpt of HXP-PI system Inference Result

System (Name)	Description	Host	Discovered System
HXP_105	Booking System	HXP on xxx2474 SLD as Bus. System	PI as Bus. System
HXP_106	Lufthansa	HXP on xxx2474 SLD as Bus. System	PI as Bus. System
HXP_107	American Airlines	HXP on xxx2474 SLD as Bus. System	PI as Bus. System
...
Interflug	Interflug	unknown	PI as Bus. Component
...
Singapore	Singapore Airlines	unknown	PI as Bus. Component

Table 6. Excerpt of HXP-PI message flow and top-level Inference Result

Top-level Connect.				
Group	Interface	Sender	Receiver	From
HXP_105 < - > HXP_106	BookOrderRequest	HXP_105	HXP_106	Config.+Runt.
	FlightSeatAvailQuery	HXP_105	HXP_106	Config.+Runt.
	FlightBookOrderConfirm	HXP_106	HXP_105	Config.+Runt.
...

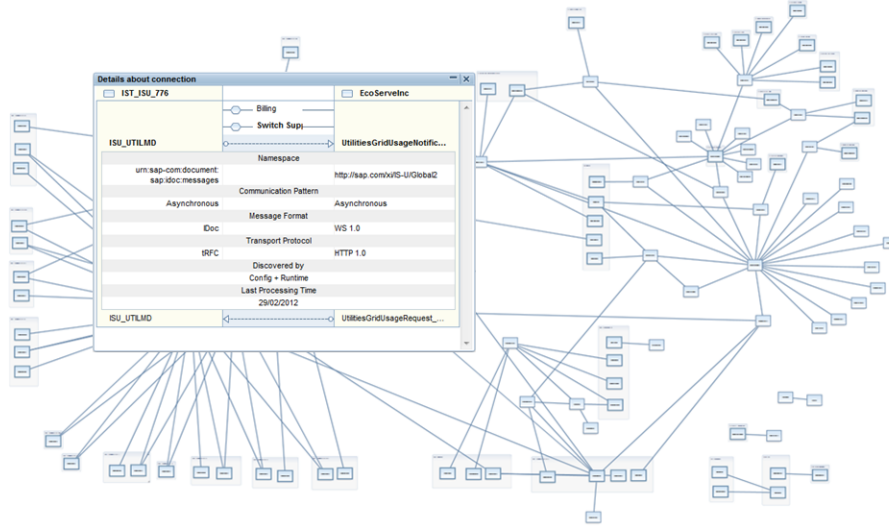


Fig. 6. Real-world customer network in a network-centric BPMN notation [19] inferred from NM raw data showing network structure and detailed view for one edge

For NM systems in general, related work is conducted in the area of Process Mining (PM) initiated by [1], which sits between computational intelligence and data mining. It has similar requirements for data discovery, conformance and enhancement with respect to NM [20], but does not work with network models and inference. PM exclusively strives to derive BPM models from process logs. Hence PM complements NM in the area of business process discovery.

Gaining insight into the network of physical and virtual nodes within enterprises is only addressed by the *Host* entity in NIM, since it is not primarily relevant for visualizing and operating integration networks. This domain is mainly addressed by the IT service management [18] and virtualization community [11], which could be considered when introducing physical entities to our meta-model.

The linked (web) data research shares similar approaches and methodologies, which have so far neglected linked data within enterprises and mainly focused on RDF-based approaches [9,10]. Applications of Datalog in the area of linked data [22,8] and semantic web [16] show that it is used in the inference domain, however not used for network inference.

8 Discussion and Future Work

In this paper we introduce a new domain for information discovery, machine learning, and network reconstruction, for which we defined a modeling and inference approach to reconstruct integration networks from NM raw data using Datalog. The network model developed specifically for the connectivity and integration domains and covers an intersection of the relevant entities, which we

derived through the analysis of several middleware systems on the market. We encoded the discovered raw data as Datalog facts to create a domain independent knowledge base and applied rule-based inference representing a multi-step network inference approach. We validated our approach on a simulated integration network and reported our experiments on applying our system to real-world enterprise networks. The evaluation shows good results with respect to the challenges like equivalence class determination, flow- and cross-middleware network reconstruction as introduced in Section 2. Although the network structure could be reconstructed very well, the discovery range should be improved to attach more integration details to the attributes of the network entity instances.

Future work will be conducted in several areas, among them the improvement of the discovery range, the inference of business process models from NM data and the correlation to integration networks as well as extensions to standard Datalog to improve the current implementation. For instance, the efficient compilation of Datalog programs to current hardware [17], distributed systems [24] or pruning with CHR [4] could guarantee more efficient Datalog processing. Since not all facts have the same certainty, we will also look into probabilistic extensions of Datalog like [26,13], which could help to express different levels of certainty with respect to network model instances. The work conducted in [3] will be considered for time aspects, which could help to prune large, outdated networks from system landscapes with historical data.

References

1. van der Aalst, W.: Process Mining: Discovery, Conformance and Enhancement of Business Processes, 2011.
2. Alvaro, P., Condie, T., Conway, N., Elmeleegy, K., Hellerstein, J.M., Sears, R.C.: BOOMAnalytics: Exploring Data-centric, Declarative Programming for the Cloud. In: EuroSys, 2010.
3. Alvaro, P., Marczak, W. R., Conway, N., Maier, D., Sears, R.: Dedalus: Datalog in Time and Space. Datalog 2.0, Oxford, 2011.
4. Campagna, D., Sarna-Starosta, B., Schrijvers, T.: Approximating Constraint Propagation in Datalog. 11th International Colloquium on Implementation of Constraint LOGic Programming Systems (CICLOPS), Lexington, KY, 2011.
5. Aref, M.: Datalog for Enterprise Applications – From Industrial Applications to Rese. Datalog 2.0 Workshop, Oxford, 2010.
6. Aref, M.: LogicBlox for Enterprise Applications. Northern California Database Day, 2011.
7. Huang, S. S., Green, T. J., Loo, B. T.: Datalog and Emerging Applications: An Interactive Tutorial. SIGMOD, 2011.
8. Abiteboul, S.: Distributed data management on the web. Datalog 2.0 Workshop, Oxford, 2010.
9. Bizer, C. Heath, T., Berners-Lee, T.: Linked Data – The Story so Far. International Journal on Semantic Web and Information Systems, Volume 5, Issue 3, p. 1–22, Elsevier, 2009.
10. Bizer, C.: The Emerging Web of Linked Data. IEEE Intelligent Systems, 24(5):87–92, 2009.

11. Chowdhury, N.M.M.K., Boutaba, R.: Network virtualization: state of the art and research challenges. *Communications Magazine*, IEEE, 2009.
12. Gallaire, H., Minker, J. (Eds.): *Logic and Data Bases*. Symposium on Logic and Data Bases, *Advances in Data Base Theory*, Plenum Press, New York, 1978.
13. Gutmann, B., Thon, I., Kimmig, A., Bruynooghe, M., De Raedt, L.: *The magic of logical inference in probabilistic programming. Theory and Practice of Logic Programming*, 2011.
14. Hellerstein, J. M.: *The Declarative Imperative – Experiences and Conjectures in Distributed Logic*. Technical Report, Berkeley, 2010.
15. Hohpe, G., Woolf, B.: *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley Longman, Amsterdam, 2003.
16. Motik, B.: *Using Datalog on the Semantic Web*. Datalog 2.0 Workshop, Oxford, 2010.
17. Neumann, T.: Efficiently Compiling Efficient Query Plans for Modern Hardware. *PVLDB* 4(9): 539-550, 2011.
18. O'Neill, P., et al.: *Topic Overview – IT Service Management*. Technical Report, Forrester Research, 2006.
19. Ritter, D., Ackermann, J., Bhatt, A., Hoffmann, F. O.: *Building a Business Graph System and Network integration Model based on BPMN*. In: 3rd International Workshop on BPMN, Luzern, 2011.
20. Ritter, D.: *From Network Mining to Large Scale Business Networks*. International Workshop on Large Scale Network Analysis (LSNA), WWW Companion, Lyon, 2012.
21. Ritter, D.: *Towards Business Network Management*. Confenis: 6th International Conference on Research and Practical Issues of Enterprise Information Systems, Ghent, 2012.
22. Polleres, A.: *Using Datalog for Rule-Based Reasoning over Web Data: Challenges and Next Steps*. Datalog 2.0 Workshop, Oxford, 2010.
23. Rodriguez, M. A., Neubauer, P.: *A Path Algebra for Multi-Relational Graphs*. International Workshop on Graph Data Management (GDM), Hannover, 2011.
24. Shaw, M., Koutris, P., Howe, B., Suciu, D.: *Optimizing Large-Scale Semi-Naive Datalog Evaluation in Hadoop*. Datalog 2.0, Vienna, 2012.
25. Ullman, J. D.: *Principles of Database and Knowledge-Base Systems Volume I*. Computer Science Press, 1988.
26. Van den Broeck, G., Thon, I., van Otterlo, M., De Raedt, L.: *DTProbLog: A decision-theoretic probabilistic Prolog*. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2010), Atlanta, 2010.